

## **DIGITAL AMPLIFICATION**

This application claims priority to U.S. Application Serial Number 09/613,826 filed July 11, 2000, which claims priority to provisional U.S. Application Serial Number 60/146,792, filed August 2, 1999.

The U.S. government retains certain rights in this invention by virtue of its support of the underlying research, supported by grants CA 43460, CA 57345, and CA 62924 from the National Institutes of Health.

### **TECHNICAL FIELD OF THE INVENTION**

This invention is related to diagnostic genetic analyses. In particular it relates to detection of genetic changes and gene expression.

### **BACKGROUND OF THE INVENTION**

In classical genetics, only mutations of the germ-line were considered important for understanding disease. With the realization that somatic mutations are the primary cause of cancer, and may also play a role in aging, new genetic principles have arisen. These discoveries have provided a wealth of new opportunities for patient management as well as for basic research into the pathogenesis of neoplasia. However, many of these opportunities hinge upon detection of a small number of mutant-containing cells among a large excess of normal cells. Examples include the detection of neoplastic cells in urine, stool, and sputum of patients with cancers of the bladder, colorectum, and lung, respectively. Such detection has been shown in some cases to be possible at a stage when the primary tumors are still curable and the patients asymptomatic. Mutant sequences from the DNA of neoplastic cells have also been found in the blood of cancer patients. The detection of residual disease in lymph nodes or surgical margins may be useful in predicting which patients

might benefit most from further therapy. From a basic research standpoint, analysis of the early effects of carcinogens is often dependent on the ability to detect small populations of mutant cells.

Because of the importance of this issue in so many settings, many useful techniques have been developed for the detection of mutations. DNA sequencing is the gold standard for the detection of germ line mutations, but is useful only when the fraction of mutated alleles is greater than ~20%.

Mutant-specific oligonucleotides can sometimes be used to detect mutations present in a minor proportion of the cells analyzed, but the signal to noise ratio distinguishing mutant and wild-type (WT) templates is variable. The use of mutant-specific primers or the digestion of polymerase chain reaction (PCR) products with specific restriction endonucleases are extremely sensitive methods for detecting such mutations, but it is difficult to quantitate the fraction of mutant molecules in the starting population with these techniques. Other innovative approaches for the detection of somatic mutations have been reviewed. A general problem with these methods is that it is difficult or impossible to independently confirm the existence of any mutations that are identified.

Thus there is a need in the art for methods for accurately and quantitatively detecting genetic sequences in mixed populations of sequences.

#### SUMMARY OF THE INVENTION

It is an object of the present invention to provide methods for determining the presence of a selected genetic sequence in a population of genetic sequences.

It is another object of the present invention to provide molecular beacon probes useful in the method of the invention.

These and other objects of the invention are achieved by providing a method for determining the presence of a selected genetic sequence in a population of genetic sequences. A biological sample comprising nucleic acid

template molecules is diluted to form a set of assay samples. The template molecules within the assay samples are amplified to form a population of amplified molecules in the assay samples of the set. The amplified molecules in the assay samples of the set are then analyzed to determine a first number of assay samples which contain the selected genetic sequence and a second number of assay samples which contain a reference genetic sequence. The first number is then compared to the second number to ascertain a ratio which reflects the composition of the biological sample.

Another embodiment of the invention is a method for determining the ratio of a selected genetic sequence in a population of genetic sequences. Template molecules within a set comprising a plurality of assay samples are amplified to form a population of amplified molecules in each of the assay samples of the set. The amplified molecules in the assay samples of the set are analyzed to determine a first number of assay samples which contain the selected genetic sequence and a second number of assay samples which contain a reference genetic sequence. At least one-fiftieth of the assay samples in the set comprise a number (N) of molecules such that  $1/N$  is larger than the ratio of selected genetic sequences to total genetic sequences required to determine the presence of the selected genetic sequence. The first number is compared to the second number to ascertain a ratio which reflects the composition of the biological sample.

According to another embodiment of the invention, a molecular beacon probe is provided. It comprises an oligonucleotide with a stem-loop structure having a photoluminescent dye at one of the 5' or 3' ends and a quenching agent at the opposite 5' or 3' end. The loop consists of 16 base pairs which has a  $T_m$  of 50-51°C. The stem consists of 4 base pairs having a sequence 5'-CACG-3'.

A second type of molecular beacon probe is provided in another embodiment. It comprises an oligonucleotide with a stem-loop structure having a photoluminescent dye at one of the 5' or 3' ends and a quenching

agent at the opposite 5' or 3' end. The loop consists of 19-20 base pairs and has a  $T_m$  of 54-56°C. The stem consists of 4 base pairs having a sequence 5'-CACG-3'.

Another embodiment provides the two types of molecular beacon probes, either mixed together or provided in a divided container as a kit.

The invention thus provides the art with the means to obtain quantitative assessments of particular DNA or RNA sequences in mixed populations of sequences using digital (binary) signals.

#### BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1. Schematic of experimental design. (A) The basic two steps involved: PCR on diluted DNA samples is followed by addition of fluorescent probes which discriminate between WT and mutant alleles and subsequent fluorometry. (B) Principle of molecular beacon analysis. In the stem-loop configuration, fluorescence from a dye at the 5' end of the oligonucleotide probe is quenched by a Dabcyl group at the 3' end. Upon hybridization to a template, the dye is separated from the quencher, resulting in increased fluorescence. Modified from Marras *et al.* (C) Oligonucleotide design. Primers F1 and R1 are used to amplify the genomic region of interest. Primer INT is used to produce single stranded DNA from the original PCR products during a subsequent asymmetric PCR step (see Materials and Methods). MB-RED is a Molecular Beacon which detects any appropriate PCR product, whether it is WT or mutant at the queried codons. MB-GREEN is a Molecular Beacon which preferentially detects the WT PCR product.

FIG. 2. Discrimination between WT and mutant PCR products by Molecular Beacons. Ten separate PCR products, each generated from ~25 genome equivalents of genomic DNA of cells containing the indicated mutations of *c-Ki-Ras*, were analyzed with the Molecular Beacon probes described in the text. Representative examples of the PCR products used for Molecular

Beacon analysis were purified and directly sequenced. In the cases with Gly12Cys and Gly12Arg mutations, contaminating non-neoplastic cells within the tumor presumably accounted for the relatively low ratios. In the cases with Gly12Ser and Gly12Asp, there were apparently two or more alleles of mutant *c-Ki-Ras* for every WT allele; both these tumors were aneuploid.

FIG. 3. Detecting Dig-PCR products with MB-RED. Specific Fluorescence Units of representative wells from an experiment employing colorectal cancer cells with Gly12Asp or Gly13Asp mutations of the *c-Ki-Ras* gene. Wells with values >10,000 are shaded yellow. Polyacrylamide gel electrophoretic analyses of the PCR products from selected wells are shown. Wells with fluorescence values <3500 had no PCR product of the correct size while wells with fluorescence values >10,000 SFU always contained PCR products of 129 bp. Non-specific products generated during the large number of cycles required for Dig-PCR did not affect the fluorescence analysis. M1 and M2 are molecular weight markers used to determine the size of fragments indicated on the left (in base pairs).

FIG. 4. Discriminating WT from mutant PCR products obtained in Dig-PCR. RED/GREEN ratios were determined from the fluorescence of MB-RED and MB-GREEN as described in Materials and Methods. The wells shown are the same as those illustrated in Fig. 3. The sequences of PCR products from the indicated wells were determined as described in Materials and Methods. The wells with RED/GREEN ratios >3.0 each contained mutant sequences while those with RED/GREEN ratios of ~1.0 contained WT sequences.

FIG. 5. Dig-PCR of DNA from a stool sample. The 384 wells used in the experiment are displayed. Those colored blue contained 25 genome equivalents of DNA from normal cells. Each of these registered positive with MB-RED and the RED/GREEN ratios were  $1.0 \pm 0.1$  (mean  $\pm$  1 standard

deviation). The wells colored yellow contained no template DNA and each was negative with MB-RED (i.e., fluorescence <3500 fluorescence units.). The other wells contained diluted DNA from the stool sample. Those registering as positive with MB-RED were colored either red or green, depending on their RED/GREEN ratios. Those registering negative with MB-RED were colored white. PCR products from the indicated wells were used for automated sequence analysis.

#### DETAILED DESCRIPTION OF THE INVENTION

The method devised by the present inventors involves separately amplifying small numbers of template molecules so that the resultant products have a proportion of the analyte sequence which is detectable by the detection means chosen. At its limit, single template molecules can be amplified so that the products are completely mutant or completely wild-type (WT). The homogeneity of these amplification products makes them trivial to distinguish through existing techniques.

The method requires analyzing a large number of amplified products simply and reliably. Techniques for such assessments were developed, with the output providing a digital readout of the fraction of mutant alleles in the analyzed population.

The biological sample is diluted to a point at which a practically usable number of the diluted samples contain a proportion of the selected genetic sequence (analyte) relative to total template molecules such that the analyzing technique being used can detect the analyte. A practically usable number of diluted samples will depend on cost of the analysis method. Typically it would be desirable that at least 1/50 of the diluted samples have a detectable proportion of analyte. At least 1/10, 1/5, 3/10, 2/5, 1/2, 3/5, 7/10, 4/5, or 9/10 of the diluted samples may have a detectable proportion of analyte. The higher the fraction of samples which will provide useful information, the more economical will be the overall assay. Over-dilution will also lead to a loss of

economy, as many samples will be analyzed and provide no signal. A particularly preferred degree of dilution is to a point where each of the assay samples has on average one-half of a template. The dilution can be performed from more concentrated samples. Alternatively, dilute sources of template nucleic acids can be used. All of the samples may contain amplifiable template molecules. Desirably each assay sample prior to amplification will contain less than a hundred or less than ten template molecules.

Digital amplification can be used to detect mutations present at relatively low levels in the samples to be analyzed. The limit of detection is defined by the number of wells that can be analyzed and the intrinsic mutation rate of the polymerase used for amplification. 384 well PCR plates are commercially available and 1536 well plates are on the horizon, theoretically allowing sensitivities for mutation detection at the ~0.1% level. It is also possible that Digital Amplification can be performed in microarray format, potentially increasing the sensitivity by another order of magnitude. This sensitivity may ultimately be limited by polymerase errors. The effective error rate in PCR as performed under our conditions was 1.1%, i.e., four out of 351 PCR products derived from WT DNA sequence appeared to contain a mutation by RED/GREEN ratio criteria. However, any individual mutation (such as a G to T transversion at the second position of codon 12 of *c-Ki-Ras*), are expected to occur in < 1 in 50 of these polymerase-generated mutants (there are at least 50 base substitutions within or surrounding codons 12 and 13 that should yield high RED/GREEN ratios). Determining the sequence of the putative mutants in the positive wells, by direct sequencing as performed here or by any of the other techniques, provides unequivocal validation of a prospective mutation: a significant fraction of the mutations found in individual wells should be identical if the mutation occurred *in vivo*. Significance can be established through rigorous statistical analysis, as positive signals should be distributed according to Poisson probabilities. Moreover, the error rate in particular Digital Amplification experiments can be precisely

determined through performance of Digital Amplification on DNA templates from normal cells.

Digital Amplification is as easily applied to RT-PCR products generated from RNA templates as it is to genomic DNA. For example, the fraction of alternatively spliced or mutant transcripts from a gene can be easily determined using photoluminescent probes specific for each of the PCR products generated. Similarly, Digital Amplification can be used to quantitate relative levels of gene expression within an RNA population. For this amplification, each well would contain primers which are used to amplify a reference transcript expressed constitutively as well as primers specific for the experimental transcript. One photoluminescent probe would then be used to detect PCR products from the reference transcript and a second photoluminescent probe used for the test transcript. The number of wells in which the test transcript is amplified divided by the number of wells in which the reference transcript is amplified provides a quantitative measure of gene expression. Another group of examples involves the investigations of allelic status when two mutations are observed upon sequence analysis of a standard DNA sample. To distinguish whether one variant is present in each allele (*vs.* both occurring in one allele), cloning of PCR products is generally performed. The approach described here would simplify the analysis by eliminating the need for cloning. Other potential applications of Digital Amplification are listed in Table 1. When the goal is the quantitation of the proportion of two relatively common alleles or transcripts rather than the detection of rare alleles, techniques such as those employing TaqMan and real time PCR provide an excellent alternative to use of molecular beacons. Advantages of real time PCR methods include their simplicity and the ability to analyze multiple samples simultaneously. However, Digital Amplification may prove useful for these applications when the expected differences are small, (*e.g.*, only ~2-fold, such as occurs with allelic imbalances.)



The ultimate utility of Digital Amplification lies in its ability to convert the intrinsically exponential nature of PCR to a linear one. It should thereby prove useful for experiments requiring the investigation of individual alleles, rare variants/mutations, or quantitative analysis of PCR products.

In one preferred embodiment each diluted sample has on average one half a template molecule. This is the same as one half of the diluted samples having one template molecule. This can be empirically determined by amplification. Either the analyte (selected genetic sequence) or the reference genetic sequence can be used for this determination. If the analysis method being used can detect analyte when present at a level of 20%, then one must dilute such that a significant number of diluted assay samples contain more than 20% of analyte. If the analysis method being used requires 100% analyte to detect, then dilution down to the single template molecule level will be required.

To achieve a dilution to approximately a single template molecule level, one can dilute such that between 0.1 and 0.9 of the assay samples yield an amplification product. More preferably the dilution will be to between 0.1 and 0.6, more preferably to between 0.3 and 0.5 of the assay samples yielding an amplification product.

The digital amplification method requires analysis of a large number of samples to get meaningful results. Preferably at least ten diluted assay samples are amplified and analyzed. More preferably at least 15, 20, 25, 30, 40, 50, 75, 100, 500, or 1000 diluted assay samples are amplified and analyzed. As in any method, the accuracy of the determination will improve as the number of samples increases, up to a point. Because a large number of samples must be analyzed, it is desirable to reduce the manipulative steps, especially sample transfer steps. Thus it is preferred that the steps of amplifying and analyzing are performed in the same receptacle. This makes the method an *in situ*, or "one-pot" method.

The number of different situations in which the digital amplification method will find application is large. Some of these are listed in Table 1. As shown in the examples, the method can be used to find a tumor mutation in a population of cells which is not purely tumor cells. As described in the examples, a probe for a particular mutation need not be used, but diminution in binding to a wild-type probe can be used as an indicator of the presence of one or more mutations. Chromosomal translocations which are characteristic of leukemias or lymphomas can be detected as a measure of the efficacy of therapy. Gene amplifications are characteristic of certain disease states. These can be measured using digital amplification. Alternatively spliced forms of a transcript can be detected and quantitated relative to other forms of the transcript using digital amplification on cDNA made from mRNA. Similarly, using cDNA made from mRNA one can determine relative levels of transcription of two different genes. One can use digital amplification to distinguish between a situation where one allele carries two mutations and one mutation is carried on each of two alleles in an individual. Allelic imbalances often result from a disease state. These can be detected using digital amplification.

Biological samples which can be used as the starting material for the analyses may be from any tissue or body sample from which DNA or mRNA can be isolated. Preferred sources include stool, blood, and lymph nodes. Preferably the biological sample is a cell-free lysate.

Molecular beacon probes according to the present invention can utilize any photoluminescent moiety as a detectable moiety. Typically these are dyes. Often these are fluorescent dyes. Photoluminescence is any process in which a material is excited by radiation such as light, is raised to an excited electronic or vibronic state, and subsequently re-emits that excitation energy as a photon of light. Such processes include fluorescence, which denotes emission accompanying descent from an excited state with paired electrons (a "singlet" state) or unpaired electrons (a "triplet" state) to a lower state with the same

multiplicity, *i.e.*, a quantum-mechanically "allowed" transition.

Photoluminescence also includes phosphorescence which denotes emission accompanying descent from an excited triplet or singlet state to a lower state of different multiplicity, *i.e.*, a quantum mechanically "forbidden" transition. Compared to "allowed" transitions, "forbidden" transitions are associated with relatively longer excited state lifetimes.

The quenching of photoluminescence may be analyzed by a variety of methods which vary primarily in terms of signal transduction. Quenching may be transduced as changes in the intensity of photoluminescence or as changes in the ratio of photoluminescence intensities at two different wavelengths, or as changes in photoluminescence lifetimes, or even as changes in the polarization (anisotropy) of photoluminescence. Skilled practitioners will recognize that instrumentation for the measurement of these varied photoluminescent responses are known. The particular ratiometric methods for the analysis of quenching in the instant examples should not be construed as limiting the invention to any particular form of signal transduction. Ratiometric measurements of photoluminescence intensity can include the measurement of changes in intensity, photoluminescence lifetimes, or even polarization (anisotropy).

Although the working examples demonstrate the use of molecular beacon probes as the means of analysis of the amplified dilution samples, other techniques can be used as well. These include sequencing, gel electrophoresis, hybridization with other types of probes, including TaqMan™ (dual-labeled fluorogenic) probes (Perkin Elmer Corp./Applied Biosystems, Foster City, Calif), pyrene-labeled probes, and other biochemical assays.

The above disclosure generally describes the present invention. A more complete understanding can be obtained by reference to the following specific examples which are provided herein for purposes of illustration only, and are not intended to limit the scope of the invention.

### EXAMPLE 1

Step 1: PCR amplifications. The optimal conditions for PCR described in this section were determined by varying the parameters described in the Results. PCR was performed in 7 ul volumes in 96 well polypropylene PCR plates (RPI). The composition of the reactions was: 67 mM Tris, pH 8.8, 16.6 mM  $\text{NH}_4\text{SO}_4$ , 6.7 mM  $\text{MgCl}_2$ , 10 mM  $\beta$ -mercaptoethanol, 1 mM dATP, 1 mM dCTP, 1 mM dGTP, 1 mM TTP, 6% DMSO, 1 uM primer F1, 1 uM primer R1, 0.05 units/ul Platinum Taq polymerase (Life Technologies, Inc.), and "one-half genome equivalent" of DNA. To determine the amount of DNA corresponding to one-half genome equivalent, DNA samples were serially diluted and tested via PCR. The amount that yielded amplification products in half the wells, usually  $\sim 1$  pg of total DNA, was defined as "one-half genome equivalent" and used in each well of subsequent Digital Amplification experiments. Fifty ul light mineral oil (Sigma M-3516) was added to each well and reactions performed in a HybAid Thermal cycler at the following temperatures: denaturation at  $94^\circ$  for one min; 60 cycles of  $94^\circ$  for 15 sec,  $55^\circ$  for 15 sec.,  $70^\circ$  for 15 seconds;  $70^\circ$  for five minutes. Reactions were read immediately or stored at room temperature for up to 36 hours before fluorescence analysis.

### EXAMPLE 2

Step 2: Fluorescence analysis. 3.5 ul of a solution with the following composition was added to each well: 67 mM Tris, pH 8.8, 16.6 mM  $\text{NH}_4\text{SO}_4$ , 6.7 mM  $\text{MgCl}_2$ , 10 mM  $\beta$ -mercaptoethanol, 1 mM dATP, 1 mM dCTP, 1 mM dGTP, 1 mM TTP, 6% DMSO, 5 uM primer INT, 1 uM MB-GREEN, 1 uM MB-RED, 0.1 units/ul Platinum Taq polymerase. The plates were centrifuged for 20 seconds at 6000 g and fluorescence read at excitation/emission wavelengths of 485 nm/530 nm for MB-GREEN and 530 nm/590 nm for MB-RED. This fluorescence in wells without template was typically 10,000 to 20,000 fluorescence "units", with about 75% emanating from the fluorometer

background and the remainder from the MB probes. The plates were then placed in a thermal cycler for asymmetric amplification at the following temperatures: 94° for one minute; 10 - 15 cycles of 94° for 15 sec, 55° for 15 sec., 70° for 15 seconds; 60° for five minutes. The plates were then incubated at room temperature for at least 20 minutes and fluorescence measured as described above. The fluorescence readings obtained were stable for several hours. Specific fluorescence was defined as the difference in fluorescence before and after the asymmetric amplification. RED/GREEN ratios were defined as the specific fluorescence of MB-RED divided by that of MB-GREEN. RED/GREEN ratios were normalized to the ratio exhibited by the positive controls (25 genome equivalents of DNA from normal cells, as defined in Materials and Methods). We found that the ability of MB probes to discriminate between WT and mutant sequences under our conditions could not be reliably determined from experiments in which they were tested by hybridization to relatively short complementary single stranded oligonucleotides, and that actual PCR products had to be used for validation.

### EXAMPLE 3

Oligonucleotides and DNA sequencing. Primer F1:

5'-CATGTTCTAATATAGTCACATTTTCA-3'; Primer R1:

5'-TCTGAATTAGCTGTATCGTCAAGG-3'; Primer INT:

5'-TAGCTGTATCGTCAAGGCAC-3'; MB-RED:

5'-Cy3-CACGGGCCTGCTGAAAATGACTGCGTG-Dabcyl-3';

MB-GREEN:

5'-Fluorescein-CACGGGAGCTGGTGGCGTAGCGTG-Dabcyl-3'.

Molecular Beacons were synthesized by Midland Scientific and other oligonucleotides were synthesized by Gene Link. All were dissolved at 50 uM in TE (10 mM Tris, pH 8.0/ 1 mM EDTA) and kept frozen and in the dark until use. PCR products were purified using QIAquick PCR purification kits (Qiagen). In the relevant experiments described in the text, 20% of the product

from single wells was used for gel electrophoresis and 40% was used for each sequencing reaction. The primer used for sequencing was 5'-CATTATTTTATTATAAGGCCTGC-3'. Sequencing was performed using fluorescently-labeled ABI Big Dye terminators and an ABI 377 automated sequencer.

#### EXAMPLE 4

Principles underlying experiment. The experiment is outlined in Fig. 1A. First, the DNA is diluted into multiwell plates so that there is, on average, one template molecule per two wells, and PCR is performed. Second, the individual wells are analyzed for the presence of PCR products of mutant and WT sequence using fluorescent probes.

As the PCR products resulting from the amplification of single template molecules should be homogeneous in sequence, a variety of standard techniques could be used to assess their presence. Fluorescent probe-based technologies, which can be performed on the PCR products "*in situ*" (i.e., in the same wells) are particularly well-suited for this application. We chose to explore the utility of one such technology, involving Molecular Beacons (MB), for this purpose. MB probes are oligonucleotides with stem-loop structures that contain a fluorescent dye at the 5' end and a quenching agent (Dabcyl) at the 3' end (Fig. 1B). The degree of quenching via fluorescence energy resonance transfer is inversely proportional to the 6<sup>th</sup> power of the distance between the Dabcyl group and the fluorescent dye. After heating and cooling, MB probes reform a stem-loop structure which quenches the fluorescent signal from the dye. If a PCR product whose sequence is complementary to the loop sequence is present during the heating/cooling cycle, hybridization of the MB to one strand of the PCR product will increase the distance between the Dabcyl and the dye, resulting in increased fluorescence.

A schematic of the oligonucleotides used for Digital Amplifications shown in Fig. 1C. Two unmodified oligonucleotides are used as primers for

the PCR reaction. Two MB probes, each labeled with a different fluorophore, are used to detect the PCR products. MB-GREEN has a loop region that is complementary to the portion of the WT PCR product that is queried for mutations. Mutations within the corresponding sequence of the PCR product should significantly impede the hybridization of it to the MB probe. MB-RED has a loop region that is complementary to a different portion of the PCR product, one not expected to be mutant. It thus should produce a signal whenever a well contains a PCR product, whether that product is WT or mutant in the region queried by MB-GREEN. Both MB probes are used together to simultaneously detect the presence of a PCR product and its mutational status.

#### Practical Considerations.

Numerous conditions were optimized to define conditions that could be reproducibly and generally applied. As outlined in Fig. 1A, the first step involves amplification from single template molecules. Most protocols for amplification from small numbers of template molecules use a nesting procedure, wherein a product resulting from one set of primers is used as template in a second reaction employing internal primers. As many applications of digital amplification are expected to require hundreds or thousands of separate amplifications, such nesting would be inconvenient and could lead to contamination problems. Hence, conditions were sought that would achieve robust amplification without nesting. The most important of these conditions involved the use of a polymerase that was activated only after heating and optimized concentrations of dNTP's, primers, buffer components, and temperature. The conditions specified in Examples 1-3 were defined after individually optimizing each of these components and proved suitable for amplification of several different human genomic DNA sequences. Though the time required for PCR was not particularly long (~2.5 hr), the number of cycles used was high and excessive compared to the number of cycles

required to amplify the “average” single template molecule. The large cycle number was necessary because the template in some wells might not begin to be amplified until several PCR cycles had been completed. The large number of cycles ensured that every well (not simply the average well) would generate a substantial and roughly equal amount of PCR product if a template molecule were present within it.

The second step in Fig 1A involves the detection of these PCR products. It was necessary to considerably modify the standard MB probe approach in order for it to function efficiently in Digital Amplification applications. Theoretically, one separate MB probe could be used to detect each specific mutation that might occur within the queried sequence. By inclusion of one MB corresponding to WT sequence and another corresponding to mutant sequence, the nature of the PCR product would be revealed. Though this strategy could obviously be used effectively in some situations, it becomes complex when several different mutations are expected to occur within the same queried sequence. For example, in the *c-Ki-Ras* gene example explored here, twelve different base substitutions resulting in missense mutations could theoretically occur within codons 12 and 13, and at least seven of these are observed in naturally-occurring human cancers. To detect all twelve mutations as well as the WT sequence with individual Molecular Beacons would require 13 different probes. Inclusion of such a large number of MB probes would not only raise the background fluorescence but would be expensive. We therefore attempted to develop a single probe that would react with WT sequences better than any mutant sequence within the queried sequence. We found that the length of the loop sequence, its melting temperature, and the length and sequence of the stem were each important in determining the efficacy of such probes. Loops ranging from 14 to 26 bases and stems ranging from 4 to 6 bases, as well as numerous sequence variations of both stems and loops, were tested during the optimization procedure. For discrimination between WT and mutant sequences (MB-GREEN probe), we



found that a 16 base pair loop, of melting temperature ( $T_m$ ) 50-51°C, and a 4 bp stem, of sequence 5'-CACG-3', were optimal. For MB-RED probes, the same stem, with a 19-20 bp loop of  $T_m$  54-56°C, proved optimal. The differences in the loop sizes and melting temperatures between MB-GREEN and MB-RED probes reflected the fact that only the GREEN probe is designed to discriminate between closely related sequences, with a shorter region of homology facilitating such discrimination.

Examples of the ratios obtained in replicate wells containing DNA templates from colorectal tumor cells with mutations of *c-Ki-Ras* are shown in Fig. 2. In this experiment, fifty copies of genomic DNA equivalents were diluted into each well prior to amplification. Each of six tested mutants yielded ratios of RED/GREEN fluorescence that were significantly in excess of the ratio obtained with DNA from normal cells (1.5 to 3.4 in the mutants compared to 1.0 in normal DNA;  $p < 0.0001$  in each case, Student's t-Test). The reproducibility of the ratios can be observed in this figure. Direct DNA sequencing of the PCR products used for fluorescence analysis showed that the RED/GREEN ratios were dependent on the relative fraction of mutant genes within the template population (Fig. 2). Thus, the DNA from cells containing one mutant *C-Ki-Ras* allele per every two WT *c-Ki-Ras* allele yielded a RED/GREEN ratio of 1.5 (Gly12Arg mutation) while the cells containing three mutant *c-Ki-Ras* alleles per WT allele exhibited a ratio of 3.4 (Gly12Asp). These data suggested that wells containing only mutant alleles (no WT) would yield ratios in excess of 3.0, with the exact value dependent on the specific mutation.

Though this mode is the most convenient for many applications, we found it useful to add the MB probes after the PCR-amplification was complete (Fig. 1). This allowed us to use a standard multiwell plate fluorometer to sequentially analyze a large number of multiwell plates containing pre-formed PCR products and bypassed the requirement for multiple real time PCR instruments. Additionally, we found that the

fluorescent signals obtained could be considerably enhanced if several cycles of asymmetric, linear amplification were performed in the presence of the MB probes. Asymmetric amplification was achieved by including an excess of a single internal primer (primer INT in Fig. 1C) at the time of addition of the MB probes.

#### EXAMPLE 5

Analysis of DNA from tumor cells. The principles and practical considerations described above was demonstrated with DNA from two colorectal cancer cell lines, one with a mutation in *c-Ki-Ras* codon 12 and the other in codon 13. Representative examples of the MB-RED fluorescence values obtained are shown in Fig. 3. There was a clear biphasic distribution, with “positive” wells yielding values in excess of 10,000 specific fluorescence units (SFU, as defined in Materials and Methods) and “negative” wells yielding values less than 3500 SFU. Gel electrophoreses of 127 such wells demonstrated that all positive wells, but no negative wells, contained PCR products of the expected size (Fig. 3). The RED/GREEN fluorescence ratios of the positive wells are shown in Fig. 4. Again, a biphasic distribution was observed. In the experiment with the tumor containing a Gly12Asp mutation, 64% of the positive wells exhibited RED/GREEN ratios in excess of 3.0 while the other 36% of the positive wells exhibited ratios ranging from 0.8 to 1.1. In the case of the tumor with the Gly13Asp mutation, 54% of the positive wells exhibited RED/GREEN ratios >3.0 while the other positive wells yielded ratios ranging from 0.9 to 1.1. The PCR products from 16 positive wells were used as sequencing templates (Fig. 4). All the wells yielding a ratio in excess of 3.0 were found to contain mutant *c-Ki-Ras* fragments of the expected sequence, while WT sequence was found in the other PCR products. The presence of homogeneous WT or mutant sequence confirmed that the amplification products were usually derived from single template molecules. The ratios of WT to mutant PCR products determined from the Digital

Amplification assay was also consistent with the fraction of mutant alleles inferred from direct sequence analysis of genomic DNA from the two tumor lines (Fig. 2).

Digital Analysis of DNA from stool. As a more practical example, we analyzed the DNA from stool specimens from colorectal cancer patients. A representative result of such an experiment is illustrated in Fig. 5. From previous analyses of stool specimens from patients whose tumors contained *c-Ki-Ras* gene mutations, we expected that 1% to 10% of the *c-Ki-Ras* genes purified from stool would be mutant. We therefore set up a 384 well Digital Amplification experiment. As positive controls, 48 of the wells contained 25 genome equivalents of DNA (defined in Materials and Methods) from normal cells. Another 48 wells served as negative controls (no DNA template added). The other 288 wells contained an appropriate dilution of stool DNA. MB-RED fluorescence indicated that 102 of these 288 experimental wells contained PCR products (mean  $\pm$  s.d. of 47,000  $\pm$  18,000 SFU) while the other 186 wells did not (2600  $\pm$  1500 SFU). The RED/GREEN ratios of the 102 positive wells suggested that five contained mutant *c-Ki-Ras* genes, with ratios ranging from 2.1 to 5.1. The other 97 wells exhibited ratios ranging from 0.7 to 1.2, identical to those observed in the positive control wells. To determine the nature of the mutant *c-Ki-Ras* genes in the five positive wells from stool, the PCR products were directly sequenced. The four wells exhibiting RED/GREEN ratios in excess of 3.0 were completely composed of mutant *c-Ki-Ras* sequence (Fig. 5B). The sequence of three of these PCR products revealed Gly12Ala mutations (GGT to GCT at codon 12), while the sequence of the fourth indicated a silent C to T transition at the third position of codon 13. This transition presumably resulted from a PCR error during the first productive cycle of amplification from a WT template. The well with a ratio of 2.1 contained a ~1:1 mix of WT and Gly12Ala mutant sequences. Thus 3.9% (4/102) of the *c-Ki-Ras* alleles present in this stool sample contained a

Gly12Ala mutation. The mutant alleles in the stool presumably arose from the colorectal cancer of the patient, as direct sequencing of PCR products generated from DNA of the cancer revealed the identical Gly12Ala mutation (not shown).